Available online at www.sciencedirect.com

SCIENCE DIRECT

Discrete Applied Mathematics 154 (2006) 650–659

DISCRETE APPLIED MATHEMATICS

www.elsevier.com/locate/dam

Non-independent randomized rounding and coloring

Benjamin Doerr[1,2]

Mathematisches Seminar, Bereich II, Christian-Albrechts-Universität zu Kiel, Christian-Albrechts-Platz 4, D-24098 Kiel, Germany

Received 30 November 2001; received in revised form 30 September 2003; accepted 6 May 2005
Available online 2 November 2005

Abstract

We propose an advanced randomized coloring algorithm for the problem of balanced colorings of hypergraphs (discrepancy problem). Instead of independently coloring the vertices with a random color, we try to use structural information about the hypergraph in the design of the random experiment by imposing suitable dependencies. This yields colorings having smaller discrepancy. We also obtain more information about the coloring, or, conversely, we may enforce the random coloring to have special properties. There are some algorithmic advantages as well.

We apply our approach to hypergraphs of d-dimensional boxes and to finite geometries. Among others results, we gain a factor $2^{d/2}$ decrease in the discrepancy of the boxes, and reduce the number of random bits needed to generate good colorings for the geometries down to $O(\sqrt{n})$ (from n). The latter also speeds up the corresponding derandomization by a factor of $\sqrt{n}$.

Keywords: Randomized algorithms; Hypergraph coloring; Discrepancy; Randomized rounding; Integer linear programming

1. Introduction and results

1.1. The discrepancy problem

The *combinatorial discrepancy problem* is to partition the vertex set of a given hypergraph into two classes in a balanced manner, i.e., in such a way that each hyperedge contains roughly the same number of vertices in each of the two partition classes.

More precisely, a *hypergraph* is a pair $\mathscr{H} = (X, \mathscr{E})$, where X is finite set and $\mathscr{E} \subseteq 2^X$. The elements of X are called *vertices*, those of $\mathscr{E}$ (*hyper*)*edges*. A partition of X into two classes is usually represented by a coloring $\chi : X \to C$ for some two-element set C. The partition then is formed by the color classes $\chi^{-1}(i)$, $i \in C$. It turns out to be useful to select -1 and $+1$ as colors. For a coloring $\chi : X \to \{-1, +1\}$ and a hyperedge $E \in \mathscr{E}$ the expression

$$\chi(E) := \sum_{x \in E} \chi(x),$$

[1] Present address: Max-Planck-Insitut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany.

[2] Partially supported (associate member) by the graduate school ‘Effiziente Algorithmen und Multiskalenmethoden’, Deutsche Forschungsgemeinschaft.

0166-218X/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2005.05.015

then counts how many of the +1-vertices of E cannot be matched by vertices colored −1. Thus $|\chi(E)|$ is a measure for the imbalance of χ with respect to E. As it is our aim to color all hyperedges simultaneously in a balanced manner, we define the discrepancy of χ with respect to $\mathscr{H}$ by

$$\mathrm{disc}(\mathscr{H}, \chi) := \max_{E \in \mathscr{E}} |\chi(E)|.$$

The discrepancy problem originates from the field of number theory (e.g. [22,18]), but due to a wide range of applications and connections it has received an increased attention in computer science and applied mathematics during the last 20 years. For reasons of brevity, we just mention uniformly distributed sets and numerical integration, computational geometry, communication complexity and image processing. We refer to the books by Matoušek [16], Chazelle [9], Kushilevitz and Nisan [13] and the paper by Asano et al. [3], respectively.

1.2. Discrepancies and integer linear programs

The discrepancy problem can be formulated as integer linear program. Since we believe that our methods can be extended to this more general context, let us briefly sketch the connection. Let $X = \{1, \ldots, n\} =: [n]$ and $\mathscr{E} = \{E_1, \ldots, E_m\}$. Then the following integer linear program (here given as a 0, 1 ILP) solves the discrepancy problem for $\mathscr{H}$:

$$\begin{array}{ll}\text{minimize} & 2\lambda \\ \text{subject to} & \displaystyle\sum_{i \in E_j} x_i - \frac{1}{2}|E_j| \leqslant \lambda, \quad j = 1, \ldots, m, \\ & \displaystyle -\sum_{i \in E_j} x_i + \frac{1}{2}|E_j| \leqslant \lambda, \quad j = 1, \ldots, m, \\ & x_i \in \{0, 1\}, \quad i = 1, \ldots, n. \\ & \lambda \geqslant 0.\end{array}$$

The problem in using the linear relaxation of this ILP is that there always exists the trivial solution $x=(x_1, \ldots, x_n)=\frac{1}{2}\mathbf{1}_n$. Therefore, a fruitful connection between solutions of the [0, 1]-relaxation and the original problem is not to be expected.

On the other hand, randomized rounding strategies for this trivial solution yield random colorings and, vice versa, generators of random colorings can be interpreted as randomized rounding strategies. Thus both problems are strongly connected. It also turns out that the tools used and the difficulties occurring in both the discrepancy problem and randomized rounding problems are very similar. Thus we believe that the methods presented in this paper might have a broader application than just the discrepancy problem. Note that due to work of Beck and Spencer [8] and Lovász et al. [14], arbitrary rounding problems can be reduced to combinatorial discrepancy problems.

1.3. Algorithmic aspects of randomized coloring and randomized rounding

Discrepancy is an NP-hard problem. For a very restricted class of discrepancy problems that are already NP-complete, we refer to Ref. [4]. Efficient algorithms finding an optimal coloring therefore are not to be expected. Indeed, very little is known about the algorithmic aspect of discrepancy. For some restrictions of the problem a nice solution exist, e.g., for hypergraphs having vertex degree at most t. Beck and Fiala [6] gave a polynomial time algorithm that computes colorings having discrepancy less than $2t$.

A common algorithmic approach for the general case (and in fact the only one known to us) are random colorings obtained by independently choosing a random color for each vertex. Via a Chernoff-bound analysis (see e.g. [2]) these colorings can be shown to have discrepancy $O(\sqrt{n \log m})$ with high probability, where as above n shall always denote the number of vertices and m the number of hyperedges.

Theorem 1. *A random coloring obtained by independently choosing a random color for each vertex has discrepancy*

$$\mathrm{disc}(\mathscr{H}, \chi) \leqslant \sqrt{2n \ln(4m)}$$

with probability at least $\frac{1}{2}$.